ELSEVIER

Linear Algebra and its Applications 304 (2000) 193–200

LINEAR ALGEBRA AND ITS APPLICATIONS

www.elsevier.com/locate/laa

Matrix rank and communication complexity

Bruno Codenotti [a,*], Gianna Del Corso [a], Giovanni Manzini [b]

[a] *Istituto di Matematica Computazionale del CNR, Via S. Maria 46, 56126 Pisa, Italy*

[b] *Università del Piemonte Orientale, Dipartimento di Scienze e Tecnologie Avanzate, Via Cavour 84, Alessandria, Italy*

Received 29 March 1999; accepted 16 September 1999

Submitted by T. Uhlig

Abstract

The rank of a matrix seems to play a role in the context of communication complexity, a framework developed to analyze basic communication requirements of computational problems. We present some issues and open problems arising in this area, and put forward a number of research subjects in linear algebra, whose investigation would shed new lights into the intriguing relationship between communication complexity and matrix rank. We also mention the related problem of the accuracy of bounds on the chromatic number of a graph given in terms of the rank of its adjacency matrix. © 2000 Elsevier Science Inc. All rights reserved.

Keywords: Matrix rank; Communication complexity; Chromatic number; Adjacency matrix; Low-rank matrices; Protocol

1. Introduction

The increasing importance of networking, telecommunication, and distributed computing has pointed out the significance of communication as a computational resource. In addition, communication plays a central role in theoretical studies: many lower bounds in complexity theory have been obtained by looking at the communication between different parts of a given computational task.

* Corresponding author.

E-mail addresses: codenotti@imc.pi.cnr.it (B. Codenotti), delcorso@imc.pi.cnr.it (G. Del Corso), manzini@al.unipmn.it (G. Manzini)

PII: S 0 0 2 4 - 3 7 9 5 (9 9) 0 0 2 2 6 - 8

We consider here a very simple model of communication, consisting of two processors (say A and B) connected by a direct link, each of them receiving its own input. Their goal is to compute a value (for simplicity we might assume this to be just one bit) which is a function of both inputs. We assume that both processors have unlimited computational power and that local computation is free, whereas we charge a unit of cost for each bit transmitted from a processor to the other one. The goal is to minimize the overall number of bits transmitted.

The central notion in communication is that of a *protocol*, which is essentially a set of rules specifying the order and the meaning of the messages sent (see, for e.g., [10,18] for more details). Associated to each protocol, there is its complexity, i.e., the number of bits transmitted in the worst case. A protocol terminates when one of the two processors, say A, knows the "answer", and the other one knows that this is the case.

There is always the so called *trivial protocol* which consists of A sending its entire input to B, so that the real challenge is to find, whenever possible, better protocols, in particular an optimal one.

A formalization of the above model can be done in terms of a matrix, called *communication matrix*, whose rows and columns are indexed by the input variables associated with processor A and B, respectively. Thus, if A and B have n and m possible inputs, respectively, we can define the $n \times m$ communication matrix $M \equiv (m_{ij})$, where the entry m_{ij} contains the value to be determined when A (resp. B) receives the ith (resp. jth) of its possible inputs.

A protocol has a very simple interpretation in the matrix setting. First of all, it determines the processor which sends the first message, say processor A. The input of A determines the first bit of information to be sent to B, and thus the protocol partitions the rows of M into two classes, where the bit transmitted by A tells B which of the two classes contains the row associated to the specific input received by A. After this, the "game" is restricted to the submatrix M_1 of M corresponding to the rows belonging to the suitable class. The next bit communicated further partitions M_1 into two classes, giving rise to a submatrix M_2, and so on. If the protocol consists of the transmission of k bits, then the matrix M_k must be the union of a submatrix of all ones and a submatrix of all zeros (a matrix with this property is called *almost homogeneous*). Indeed A knows the output bit when looking at M_k if its row in M_k has constant entries; furthermore this must be true for all rows, otherwise B would not know that the protocol has ended.

A slight variation of the above definition of protocol consists of ending it in the presence of a monochromatic (either all zeros or all ones) matrix, instead of an almost homogeneous one.

It is immediate to notice that a protocol can be analyzed in terms of ranks of the submatrices detected as the rounds proceed. At each step the maximum rank of submatrices decreases by at most a factor of 2 so that one obtains the following easy but important result.

Theorem 1 [10]. *Let $c(M)$ be the communication complexity associated to the* $\{0, 1\}$ *matrix M, i.e., the minimum number of bits that must be transmitted in any protocol associated with M. Then* $c(M) \geqslant \log_2(\mathrm{rank}(M))$.

From Theorem 1 we can immediately derive the following corollary.

Corollary 2. *If M has full rank, then the trivial protocol, which consists of transmitting a row or column index is optimal.*

It is worthwhile to mention that the rank of the communication matrix is an upper bound on communication complexity. In fact, it is not difficult to prove the following theorem.

Theorem 3 [10]. $c(M) \leqslant \mathrm{rank}(M)$.

Theorems 1 and 3 show that communication complexity can be bounded in terms of rank, and has focussed much of the current research on understanding whether the upper bound can be sharpened. In particular, the following conjecture has been raised.

Conjecture 1 (*see* [9,16]). $c(M) \leqslant [\log_2(\mathrm{rank}(M))]^c$ for a positive constant c.

Conjecture 1 is closely related with the accuracy of bounds on the chromatic number of a graph obtained in terms of the rank of its adjacency matrix. Indeed, Lovázs and Saks proved in [9] that Conjecture 1 is true if and only if there exists a constant c such that the chromatic number of any graph G does not exceed $\exp(\log^c(\mathrm{rank}(A_G)))$, where A_G denotes the adjacency matrix of G.

2. Brief history

Interest in communication complexity within the theoretical computer science community started with research activities on the theory of distributed computing, and in particular, on area-time tradeoffs for VLSI circuits. Indeed the paper by Mehlhorn and Schmidt [10], as well as several other seminal contributions (see, e.g., [1,18,19]), were mainly motivated by the above research interests.

Very soon, during the 1980s, the research community became aware that communication also plays a central role in computational complexity per se; indeed communication was recognized as a central issue for analyzing how different parts of a computation must interact, and how the nature of such interaction could be related to complexity [13].

This led to identifying a number of nontrivial research problems of an algebraic flavour whose solution is still open. Among those, there is the problem of the gap between the rank of the communication matrix and its communication complexity.

This problem has two facets: on the one side, it amounts to looking for explicit matrices with the largest possible gap between rank and communication complexity (see Section 3); on the other side, to finding general properties of low rank matrices which could possibly lead to a proof of Conjecture 1 (see Section 3).

The related problem of the gap between chromatic number and rank originated from a conjecture of van Nuffelen [12], which was then rediscovered by Fajtlowicz's computer program "Graffiti" [4]. The conjecture stated that the chromatic number cannot exceed the rank of its adjacency matrix. This was proven false by Alon and Seymour [2], a superlinear gap was then found by Razborov [17], and a larger gap was provided by Nisan and Wigderson [11]. The complementary issue of upper bounds on the chromatic number in terms of rank has been investigated in [6].

3. The gap

Conjecture 1, as well as the related conjecture concerning the chromatic number, has stimulated research activities oriented towards finding matrices whose communication complexity is "much larger" than the logarithm of the rank [2,11,16,17]. The best known separation result is due to Nisan and Wigderson who have constructed an infinite family of matrices such that $c(M) \geqslant [\log_2(\mathrm{rank}(M))]^{\alpha}$, with $\alpha = \log_2 3 \sim 1.58$. More precisely, they described a $2^n \times 2^n$ matrix M such that $c(M) = \Omega(n)$ and $\log_2(\mathrm{rank}(M)) = \mathrm{O}(n^{\log_3 2})$ [11, Theorem 1].

In the following we sketch their construction.

Let $E : \{0,1\}^3 \to \{0,1\}$ denote the Boolean function which evaluates to 1 iff either one or two of its inputs are equal to 1. Note that E can be expressed as a degree 2 polynomial, i.e., $E(z_1, z_2, z_3) = z_1 + z_2 + z_3 - z_1 z_2 - z_1 z_3 - z_2 z_3$.

We now define a function $E_k : \{0,1\}^{3^k} \to \{0,1\}$ recursively, by setting $E_0(z) = z$, and $E_k(\cdot) = E(E_{k-1}(\cdot), E_{k-1}(\cdot), E_{k-1}(\cdot))$, where the domain of each E_{k-1} is a different set of 3^{k-1} Boolean variables. Let $n = 3^k$. To the function E_k we associate the $2^n \times 2^n$ matrix M defined as

$$M_{ij} = E_k(i_1 j_1, i_2 j_2, \ldots, i_n j_n),$$

where $i_1 i_2 \cdots i_n$ and $j_1 j_2 \cdots j_n$ denote the binary representations of i and j, respectively.

One can easily prove by induction that E_k is *fully sensitive* at $\vec{0}$, i.e., that $E_k(\vec{0}) = 0$, whereas $E_k(\vec{u}) = 1$ for any input vector $\vec{u}$ with exactly one bit equal to 1. Let P_t be the set of positions containing a 1 in the binary representation of the integer t. Since E_k is fully sensitive, then $M_{ij} = 0$ if $P_i \cap P_j = \emptyset$, and $M_{ij} = 1$ if $P_i \cap P_j$ contains a single element. The value taken by the other entries of M is irrelevant.

This shows that the matrix M "evaluates" to 0 whenever two n-element sets have an empty intersection and to 1 if their intersection contains a single element. Since every function with this property is known to have communication complexity $\Omega(n)$ (see for example [5]), we obtain the lower bound $c(M) = \Omega(n)$.

In order to prove the required upper bound on $\log_2(\mathrm{rank}(M))$, we notice that the function $E_k(z_1, \ldots, z_n)$ can be expressed as a polynomial of degree 2^k, i.e.,

$$E_k(z_1, \ldots, z_n) = \sum_S \alpha_S \prod_{k \in S} z_k,$$

where the summation is over all subsets of $\{1, \ldots, n\}$ of size at most 2^k. For each subset S, we define the matrix $M_{ij}^{(S)} = \prod_{k \in S} i_k j_k$, where $i_1 i_2 \cdots i_n$ and $j_1 j_2 \cdots j_n$ are the binary representations of i and j. Clearly $M = \sum_S \alpha_S M^{(S)}$, and each $M^{(S)}$ has rank 1. Hence, the rank of M is bounded from above by the number of nonzero monomials in the representation of E_k, which is less than 6^{2^k-1}. This yields the bound $\log_2(\mathrm{rank}(M)) = \mathrm{O}(2^k) = \mathrm{O}(n^{\log_3 2})$ as claimed.

Note that if we start with a fully sensitive Boolean function on n variables with degree d, the above construction yields a family of matrices for which $c(M) \geqslant [\log_2(\mathrm{rank}(M))]^\alpha$, with $\alpha = \log_d n$. However, since a fully sensitive function on n variables must have degree at least $\sqrt{n}/2$, this approach can only produce a gap with an exponent at most 2.

It is interesting to notice that the matrix M satisfies a recursive formula based on the Kronecker product [3]. A similar recurrence has also been proven for another class of matrices characterized by a nonconstant gap between communication complexity and logarithm of the rank, namely for the matrices introduced in [16] which satisfy $c(M) = \Omega(\log_2(\mathrm{rank}(M)) \log_2 \log_2 \log_2(\mathrm{rank}(M)))$.

4. Properties of low rank matrices

We have already observed that nontrivial communication protocols can only arise for low rank matrices. In these cases, the truth of Conjecture 1 would intuitively imply that a protocol could end quickly by taking advantage of certain general properties of submatrices of low rank matrices. Indeed we will shortly see that one of these properties is the existence, in any low rank matrix, of a very large monochromatic submatrix.

Let $|B|$ denote the size of a matrix B, i.e., the number of rows times the number of columns. Let $\mathrm{mono}(M)$ be the maximum value attained by the ratio $|A|/|M|$, taken over all monochromatic submatrices A of M.

An easy relation between $\mathrm{mono}(M)$ and $c(M)$ follows by observing that, after $c(M)$ steps, a protocol induces a partition of M into at most $2^{c(M)}$ monochromatic submatrices. Since at least one of these submatrices must have size greater than $|M|/2^{c(M)}$, then $\mathrm{mono}(M) \geqslant 2^{-c(M)}$.

A possible connection between $\mathrm{mono}(M)$ and $\mathrm{rank}(M)$ is provided by Conjecture 2 below, due to Nisan and Wigderson [11], who also proved its equivalence to Conjecture 1.

Conjecture 2. $-\log_2(\mathrm{mono}(M)) \leqslant [\log_2(\mathrm{rank}(M))]^\alpha$ for some positive constant α.

As a consequence, a viable way to prove Conjecture 1 could consist of showing that every low rank matrix contains a suitably large monochromatic submatrix.

Aware of the difficulty of proving significant bounds on $\mathrm{mono}(M)$ in terms or rank, Nisan and Wigderson introduced the more tractable notion of *discrepancy*, which is still somehow related to rank and communication complexity.

Let A be a submatrix of M and let a_0 and a_1 denote the number of entries equal to 0 and to 1 in A, respectively. The discrepancy of A is defined as $\delta(A) = |a_0 - a_1|/|M|$, and $\mathrm{disc}(M)$ is the maximum value attained by $\delta(A)$, taken over all submatrices A.

Note that if A is a monochromatic submatrix, then $\delta(A) = |A|/|M|$, from which $\mathrm{disc}(M) \geqslant \mathrm{mono}(M)$ follows.

The relation between discrepancy and communication complexity can be intuitively understood in the context of probabilistic protocols, where one could imagine a protocol ending in the presence of a submatrix with large discrepancy, outputting the most likely answer. The connection with rank is explained by the following theorem.

Theorem 4 [11]. $\mathrm{disc}(M) \geqslant c \,\mathrm{rank}(M)^{-3/2}$ *for some positive constant c.*

We now present the idea behind the proof. Instead of $\{0, 1\}$ matrices, we consider $\{-1, +1\}$ matrices. This does not change the discrepancy, whereas the rank increases or decreases by at most 1. The advantage is that now, for a submatrix N, $\delta(N)$ is simply the sum of its entries divided by $|M|$. When M has low rank, we wish to find a submatrix of high discrepancy, or equivalently two $\{0, 1\}$ vectors $\mathbf{x}$ and $\mathbf{y}$ such that the absolute value of $\mathbf{x}^{\mathrm{T}} M \mathbf{y}$ is large. The nonzero entries of $\mathbf{x}$ and $\mathbf{y}$ identify a submatrix B of high discrepancy. The existence of such a B is implied by the following two results, which can be proved by the manipulation of L_2 and L_∞ norms and by using known trace, norms, and rank inequalities.

(1) For any $\{-1, +1\}$ matrix A of size n, there exist vectors $\mathbf{u}$, $\mathbf{v}$, with $\|\mathbf{u}\|_\infty \leqslant 1$ and $\|\mathbf{v}\|_\infty \leqslant 1$, such that $\mathbf{u}^{\mathrm{T}} A \mathbf{v} \geqslant n^2/(16\,(\mathrm{rank}(A))^{3/2})$.

(2) For any $\{-1, +1\}$ matrix A of size n and for any pair of vectors $\mathbf{u}$ and $\mathbf{v}$ such that $\|\mathbf{u}\|_\infty \leqslant 1$ and $\|\mathbf{v}\|_\infty \leqslant 1$, there exists a submatrix B of A, such that $\delta(B) \geqslant \mathbf{u}^{\mathrm{T}} A \mathbf{v}/(4n^2)$.

Combining (1) and (2), it is easy to see that M must contain a submatrix B such that $\delta(B) \geqslant 1/(64\,\mathrm{rank}(M)^{3/2})$.

The lower bound given in Theorem 4 is sharp, since there are infinitely many matrices of a given rank such that $\mathrm{disc}(M) \leqslant 1/\mathrm{rank}(M)$. Summarizing, Theorem 4 successfully deals with $\mathrm{disc}(M)$ (which is $\geqslant \mathrm{mono}(M)$), thus raising the issue of possible gaps between $\mathrm{disc}(M)$ and $\mathrm{mono}(M)$.

5. Research issues

The main open question of algebraic interest is that of quantifying the gap between rank and communication complexity, especially in connection with Conjecture 1. If

Conjecture 1 is true, then there is a strong correlation between the logarithm of the rank of the communication matrix and its communication complexity. On the other hand, if it does not hold, then there are matrices whose high communication complexity depends on properties only loosely correlated with the rank. An interesting research topic along the above lines is the following.

Problem 1. Identify algebraic notions responsible for different communication complexities of $\{0, 1\}$ matrices with the same rank.

Other important questions arise from the subject treated in Section 4, where we have insisted on properties of low rank matrices which may be relevant for communication complexity.

An interesting observation by Nisan and Wigderson is that instead of proving the existence of a large monochromatic (rank 1) submatrix, one could equivalently prove the existence of a large submatrix with rank $(1 - \varepsilon)\mathrm{rank}(M)$, for an arbitrary constant $\varepsilon > 0$, and still obtain the same result, i.e., a proof of Conjecture 2.

Thus the following questions seem to be of major importance.

Problem 2. Let M be a $\{0, 1\}$ matrix of rank r.
1. Find the largest submatrix of M of rank 1;
2. Find the largest submatrix of M of rank αr, $\alpha < 1$.

Problem 2 seems to have inspired the work by Kotlov and Lovász [7], who analyzed the size of submatrices of rank less than r. They proved that no two columns of M can coincide on a number of positions exceeding the size of the largest submatrix A of rank less than r, and that, if such an A has identical rows then its rank is $r - 2$. By using the above properties, they also show that if M does not have identical rows, then its size is $O(2^{r/2})$.

Another way of looking at the above issues is the following.

Problem 3. Find the largest possible gap between $\mathrm{disc}(M)$ and $\mathrm{mono}(M)$.

A full understanding of Problems 2 and 3 would lead to significant advances in the theory of communication complexity, and also in other branches of theory of computing.

Proving either partial or narrower results could help to make progress. For instance, partial results on Problems 2 and 3 could be interesting, as well as any result implying a closer link between discrepancy and rank, e.g., any improvement over the $3/2$ exponent in the bound of Theorem 4.

It is finally worthwhile to mention the existence of different algebraic notions (variation rank, contact rank, and tensor rank) which have been used in connection with communication complexity and could provide additional insights (see, e.g., [8,14,15]).

References

[1] A.V. Aho, J.D. Ullman, M. Yannakakis, On the notion of information transfer in VLSI circuits, in: Proceedings 15th Annual ACM Symposium on the Theory of Computing, 1983, pp. 133–139.
[2] N. Alon, P. Seymour, A counterexample to the rank-coloring conjecture, J. Graph Theory 13 (1989) 523–525.
[3] B. Codenotti, G. Manzini, L. Margara, Algebraic techniques in communication complexity, Inform. Process. Lett. 56, (1995) 191–195.
[4] S. Fajtlowicz, On conjectures of graffiti, Discrete Math. 72 (1988) 113–118.
[5] B. Kalyanasundaram, G. Schnitger, The probabilistic communication complexity of set intersection, in: Second Structure in Complexity Theory Conference, 1987, pp. 41–49.
[6] A. Kotlov, Rank and chromatic number of a graph, J. Graph Theory 26 (1997) 1–8.
[7] A. Kotlov, L. Lovász, The rank and size of graphs, J. Graph Theory 23 (1) (1996) 185–189.
[8] M. Krause, S. Waack, Variation ranks of communication matrices and lower bounds for depth two circuits having symmetric gates with unbounded fan-in, in: Proceedings of the 32nd Annual IEEE Symposium on Foundations of Computer Science, 1991, pp. 777–782.
[9] L. Lovász, M. Saks, Lattices, Möbius functions and communication complexity, in: Proceedings of the 29th Annual IEEE Symposium on Foundations of Computer Science, 1988, pp. 81–90.
[10] K. Mehlhorn, E.M. Schmidt, Las Vegas is better than determinism in VLSI and distributed computing, in: Proceedings 14th Annual ACM Symposium on the Theory of Computing, 1982, pp. 330–337.
[11] N. Nisan, A. Wigderson, On rank vs. communication complexity, Combinatorica 15 (1995) (Preliminary version in: Proceedings of the 35th Annual IEEE Symposium on Foundations of Computer Science, 1994, pp. 831–836).
[12] C. van Nuffelen, A bound for the chromatic number of a graph, Amer. Math. Monthly 83 (1976) 265–266.
[13] P. Pudlák, Communication in bounded depth circuits, Combinatorica 14 (2) (1994) 203–216.
[14] P. Pudlák, V. Rödl, Modified ranks of tensors and the size of circuits, in: Proceedings 25th Annual ACM Symposium on the Theory of Computing, 1993, pp. 523–531.
[15] P. Pudlák, V. Rödl, J. Sgall, Boolean circuits, tensor ranks and communication complexity, SIAM J. Computing 26 (3) (1997) 605–633.
[16] R. Raz, B. Spiker, On the log-rank conjecture in communication complexity, Combinatorica 15 (1995) (Preliminary version in: Proceedings of the 34th Annual IEEE Symposium on Foundations of Computer Science, 1993, pp. 168–176).
[17] A. Razborov, The gap between the chromatic number of a graph and the rank of its adjacency matrix is superlinear, Discrete Math. 108 (1982) 393–396.
[18] A.C. Yao, Some complexity questions related to distributive computing, in: Proceedings of the 11th Annual ACM Symposium on the Theory of Computing, 1979, pp. 209–213.
[19] A.C. Yao, The entropic limitations of VLSI computations, in: Proceedings of the 13th Annual ACM Symposium on the Theory of Computing, 1981, pp. 308–311.